

СЕМАНТИЧЕСКИЕ ТЕХНОЛОГИИ В ИНТЕГРАЦИИ ДАННЫХ ФИНАНСОВЫХ ОРГАНИЗАЦИЙ

Петрова Г.Г., Тузовский А. Ф.

Томский политехнический университет
ggp_pgg@mail.ru

Введение

Современные финансовые организации используют большое количество данных, содержащихся в своих информационных системах. Эти данные хранятся в многочисленных источниках, использующих различные технологии и имеющих различные логические структуры. Для повышения эффективности использования всех данных требуется предоставить единый интерфейс для доступа к совокупности неоднородных независимых источников данных – созданию системы интеграции данных. Такая система должна освободить пользователей от необходимости знать местоположение всех источников данных, их свойства и способы доступа к ним. Интегрируемыми источники данных могут быть различные базы данных, электронные таблицы, веб-сервисы. Доступ к данным из различных источников означает, что совокупность данных из множества независимых источников будет представлена в терминах единой модели данных. Одним из подходов, позволяющим выполнить такое представление структуры данных и реализовать доступ к ним, являются семантические технологии.

Системы интеграции данных

Первые шаги в области интеграции информационных ресурсов относятся к середине 70-х гг., когда разработки распределенных систем баз данных позволили сформировать представления о многоуровневой архитектуре систем баз данных, о моделях данных как инструменте моделирования реальности и об отображении моделей данных. Речь при этом шла, главным образом, о поддержке глобальной схемы для совокупности локальных баз данных, функционирующих в разных узлах сети под управлением СУБД, которые поддерживают одну и ту же или, в общем случае, разные модели данных. Позднее несколько более общая форма этой задачи была связана с созданием мультимедийных баз данных, хранилищ данных, различных репозиториях информационных ресурсов, а также веб-приложений. В последние годы в широко развернувшихся во многих странах разработках электронных библиотек (Digital Libraries) [1] проблемы интеграции неоднородных данных стали играть ключевую роль, причем возникает также задача интеграции текстовых информационных ресурсов из различных независимых источников [2].

Известные технологические подходы к интеграции информационных систем в основном сводятся к следующим способам:

1. "Точка-точка". Может быть реализована различными техническими средствами, но суть метода сводится к прямой передаче информации из одной системы в другую. Например, организация периодической выгрузки данных из одной ИС в какой-либо файл, и загрузка этого файла в другие ИС. Операции загрузки и выгрузки выполняются специально написанными программными модулями, или специализированными приложениями для обмена данными. Другой вариант - организация прямого доступа базы данных одной ИС к базе других ИС, копирование данных на уровне базы данных. В этом случае преобразование структуры переносимых данных осуществляется программными процедурами, созданными при помощи инструментальных средств самой базы данных.

2. SOA (сервисно-ориентированная архитектура). При таком варианте интеграции каждая информационная система предлагает другим определенный набор сервисов, связанных со спецификой хранящихся в ней данных. Каждый сервис позволяет выполнить какую-то определенную операцию: например, узнать реквизиты клиента, добавить контактную персону, и т.д.

Перечисленные технологические подходы не стандартизированы каким-либо образом, и используются в широком круге информационных систем.

Недостатки этих подходов состоят в следующем:

- оба подхода зависят от структуры данных информационной системы;

- при одновременном изменении одной и той же информации в системе интеграции может возникнуть конфликт;

- при интеграции трех и более систем повышается риск возникновения сбоев;

- в связи с тем, что обмен информацией происходит не в режиме реального времени, пользователь может работать с уже устаревшими данными.

Семантическая интеграция подразумевает, что информационные системы обмениваются данными в такой форме, которая полностью абстрагирована от внутренней структуры каждой из них. Информация записывается в виде "высказываний" (на языке RDF) при помощи определенного "словаря" (онтологии, описываемой на языке

OWL); системы отправляют и получают эти сообщения по различным правилам. При этом технически информация может передаваться путем записи в файлы, или через веб-сервисы.

Преимущества использования семантических технологий состоят в следующем:

- семантические технологии позволяют осуществлять интеграцию нескольких (двух и более) информационных систем, независимо от их структуры с низкими трудозатратами на написание программного кода;
- при изменении структуры информации в одной из систем обмен не будет нарушен;
- значительно упрощается формирование обобщенных аналитических срезов данных, формирующихся на основании объединения информации из нескольких систем.

Все перечисленные преимущества существенно снижают трудовые и материальные затраты на внедрение и поддержку средств интеграции данных, позволяют обеспечить высокую оперативность передачи информации между системами [3].

Интеграция данных с использованием семантических технологий на примере информационных систем предприятия

Рассмотрим концептуальную модель интеграции информационных систем финансового предприятия: базы данных MySQL и базы данных 1С. Предприятие работает в области кредитования малого и среднего бизнеса, а также привлекает средства инвесторов. База данных MySQL обеспечивает функционал финансового планирования (планируемые графики платежей заемщиков и планируемые графики выплат инвесторам), база данных 1С обеспечивает функционал финансового учёта (фактические поступления средств от заемщиков и фактические выплаты инвесторам). Для получения полной картины финансового состояния предприятия необходимо интегрировать вышеперечисленные информационные системы.

При создании системы интеграции необходимо решить следующий ряд задач:

- Разработать архитектуру системы интеграции данных. Архитектура системы интеграции данных для финансового предприятия представлена на рисунке 1;



Рис. 6. Архитектура системы интеграции

- Создать интегрирующую модель данных, являющейся основой единого пользовательского интерфейса в системе интеграции;
- Разработать методы отображения моделей данных и построения отображений в интегрирующую модель для баз данных MySQL и 1С;
- Интегрировать метаданные, используемые в базах данных MySQL и 1С;
- Определить пути преодоления неоднородности источников данных;
- Разработать механизмы семантической интеграции источников данных.

В качестве интегрирующей модели данных предлагаем использовать онтологию FIBO (Financial Industry Business Ontology) [4] – это модульная формальная модель понятий, описываемых терминами финансовой отрасли, которые используются в официальных документах финансовых организаций. Данная онтология разрабатывается под управлением консорциума OMG, занимающегося разработкой и продвижением объектно-ориентированных технологий и стандартов. Она называется «Концептуальной моделью финансового бизнеса», что отличает ее от моделей и описаний данных или их реализаций в информационных технологиях. FIBO предназначена для использования финансовыми бизнес-аналитиками и другими участниками сферы финансов.

Предлагаемое решение может быть использовано не только конкретным предприятием, указанным в примере, но и другими финансовыми компаниями: банками, микрофинансовыми организациями.

Заключение

Семантические технологии являются одним из перспективных инструментов интеграции данных, позволяющим исследовать финансовые данные в масштабах предприятия и осуществляющим интеграцию данных с учетом бизнес-требований.

Список использованных источников

1. Бездушный А.Н., Жижченко А.Б., Кулагин М.В., Серебряков В.А. Интегрированная система информационных ресурсов РАН и технология разработки цифровых библиотек. Программирование, 2000, 4, с. 3-14.
2. Когаловский М.Р. Методы интеграции данных в информационных системах. Москва, 2010.
3. Бизнес семантика [Электронный ресурс] – 2015 – Режим доступа: <http://www.business-semantic.ru/>, свободный.
4. Object Management Group [Электронный ресурс] – 2015 – Режим доступа: <http://www.omg.org/>, свободный.